

# Responsabilidad moral respecto a robots dotados con moralidad artificial heredada culturalmente mediante neuro-memes

*Dante Giovanni Sterpin\**

## Resumen

Bajo la premisa de que todo artefacto tecnológico puede implicar un efecto moralmente significativo para el ser humano, en particular por la función para la cual fue diseñado, los robots que están destinados a interactuar socialmente con personas de la tercera edad, por ejemplo, están pensados con miras a que, de algún modo, puedan mejorar la calidad de vida del adulto mayor. Así, dichos artefactos son agentes morales que “inconscientemente” implican bienestar. Sin embargo, la tendencia a incrementar su autonomía perceptiva, motriz y comunicativa, para servir como colaboradores o compañía alternativa, también incrementa el riesgo de ser empleados de maneras impredeciblemente negativas. Esto acarrea la responsabilidad de reducir sus efectos inmorales, dotándolos de autonomía moral y educándolos para ser responsables de sus propios actos.

**Palabras clave:** moralidad artificial, neuro-memética, robot autónomo

## Abstract

Under the premise that any technical artefact can have a morally significant effect on a human being, particularly because of the function which it was designed for, robots that are designed to socially interact with the elderly –for example–, are conceived as a service that, in some way, can improve their life. Thus, such artefacts are moral agents that “unconsciously” imply well-being. However, the tendency to increase its perception, mobility and communication autonomy, to act as collaborators or alternative company, also increases the risk of being employed in unpredictably negative ways. So we are responsible for reducing their immoral effects, equipping them with moral autonomy as well, and educating them to be responsible for their own actions too.

**Keywords:** autonomous robot, artificial morality, neuro-memetics.

\* Ingeniero Electrónico, Universidad Santo Tomás. Especialista en Docencia Universitaria, Universidad Militar Nueva Granada. Estudiante de Maestría en Ingeniería de Sistemas y Computación, Pontificia Universidad Javeriana. dante\_sterpin@javeriana.edu.co

## Introducción

Para el desarrollo de robots autónomos que emplean inteligencia artificial, sobre todo los denominados *humanoides* en interacción social con seres humanos vulnerables (Tapus, Mataric y Scassellati, 2007), se considera en su diseño el tipo de valor que garantiza buen desempeño motor, con capacidad visual para el reconocimiento de objetos y rostros, y auditiva para recibir órdenes habladas; así como capacidad de comunicación social. Sin embargo, considerar un buen comportamiento, es decir, que sea moralmente aceptable, es un valor que va más allá del buen funcionamiento. Así, se considera factible y necesaria la incorporación de moral artificial en los robots autónomos (Allen, Wallach y Smith, 2006), con respecto a lo cual, ya se han considerado algunos modelos computacionales basados en técnicas evolutivas (Wallach, 2008).

En ese contexto, puede considerarse apropiado emplear en el diseño de robots con moral artificial un modelo computacional fundamentado en dinámicas evolutivas de carácter cultural, puesto que la emergencia, propagación y variación de comportamientos, ideas, creencias y normatividades morales son dinámicas propias de la cultura. Al respecto, se tiene disponible la hipótesis del meme como posible replicador base en la evolución cultural (Dawkins, 1976) y a la neuro-memética, como una especificación del sustrato material que le da soporte al replicador (Aunger, 2002). Actualmente hay algunos modelos informáticos sobre la evolución cultural basada en memes, entre los cuales está el autómatas memético (Chen, Ong, Lim, y Tan, 2011); el modelo de la herencia dual (Marriott y Chebib, 2016) y el perceptrón auto-supervisado (Sterpin, 2011).

Ahora bien, de la evolución cultural hay una gran cantidad de productos que van desde ideas

hasta artefactos, incluyendo tradiciones religiosas y las legislaciones que en general establecen las normas del comportamiento moral. También se podría sopesar la hipótesis de que heredemos una gramática moral que co-evolucionó con el cerebro, por lo que nuestros juicios morales operarían, en principio, de manera inconsciente e intuitiva, y luego se racionalizan. Esto implica un desarrollo evolutivo que pudo fundamentarse en ciertos instintos sociales y la consecuente necesidad de altruismo recíproco (Hauser, 2006). Por esto, la moral podría fundamentarse biológicamente y reforzarse culturalmente.

Por otro lado, la simulación mental es uno de los aspectos fundamentales de la evolución cultural (Gabora, 1995), mientras que la empatía puede ser la emoción sobre la cual se fundamenta nuestra forma de “leer la mente” ajena, como una suerte de simulación mental, es decir, mediante imaginarios con los cuales uno puede ubicarse virtualmente en la situación del otro, o “ponerse en sus zapatos”, para inferir sobre su estado mental a partir de su comportamiento observable (Krueger, 2013). Emular la empatía es uno de los grandes retos en el desarrollo de los robots para la asistencia social (Tapus et al., 2007). Como a la par, dicha emoción puede verse como un fundamento en el desarrollo moral (Stueber, 2013), sus características deben considerarse en el diseño del modelo computacional que, además de darle la capacidad de ser empático al interactuar con personas, le dé la capacidad de llegar a comportarse según los estándares de la sociedad en donde se requiera su servicio.

El rol característico de la empatía en la comprensión del otro está muy ceñido al contexto del aprendizaje imitativo con base en las “neuronas espejo” (Krueger, 2013; Stueber, 2013). Sin

embargo, desde la neuro-memética, la imitación no parece ser una buena perspectiva para abordar la propagación de características culturales (Aunger, 2002), pues lo que parece imitación puede deberse a la replicación de neuro-memes, razón por la cual esta perspectiva, y su modelo (Sterpin, 2011), fundamentan la presente reflexión.

## Situación Hipotética

En una situación en la que un ser humano asesina a otro, se sabe que el homicida debe ser juzgado y castigado según los paradigmas morales del contexto cultural implicado. En contraste, dado el caso de que un robot asesine a un ser humano, se juzgaría y castigaría, no al robot mismo, sino a su diseñador. Pero, hipotéticamente, si ese robot tuviese la misma autonomía moral de un ser humano, ¿el diseñador sigue siendo culpable de ello?

## Robots humanistas

Los artefactos dotados con técnicas de la llamada “inteligencia artificial”, que incrementan su nivel de autonomía en cuanto a tomar decisiones en la interacción social con seres humanos, dejan de ser artefactos neutrales en la medida en que esas acciones pueden implicar efectos moralmente significativos (Vermaas, Kroes, van de Poel, Franssen, y Houkes, 2011). Imagine ahora esta segunda situación hipotética: un homicidio perpetrado por un robot asistencial que decidió desconectar el soporte vital de un adulto mayor, amputado de brazos y piernas, pues este le argumentó verbalmente que sufría y nadie respetaba

Este artículo presenta una reflexión sobre la responsabilidad implicada en emplear neuro-memes artificiales, para emular la herencia cultural de comportamientos morales. Sin embargo, comprender la naturaleza de la moral no es una tarea trivial y quizás exija confrontar ideas fuertemente atadas en ciertas creencias religiosas. Sin embargo, acá tan solo se indaga sobre la necesidad de robots con moralidad artificial y se consideran algunos aspectos técnicos y filosóficos que viabilizan su desarrollo.

En este caso hipotético se asume que el diseñador incorporó la responsabilidad de estudiar la naturaleza moral para hacer que dicho robot tenga “autonomía moral”. De esa manera, en este artículo se plantea que el diseñador ya no tiene culpa frente a las acciones que decida el robot, pues este último debió construir su discernimiento moral por inmersión en una sociedad que debe educarlo en ese sentido, de forma semejante a como se educa a los seres humanos.

su deseo de ponerle fin a su vida. Así, se asume que el robot es capaz de procesar el lenguaje y, de alguna manera, entender lo que se le dice. Acudiendo a nuestra capacidad de imaginar experiencias ajenas como si fueran propias, por un breve momento podemos visualizarnos en esa misma situación, preguntándonos, quizás, si nos atreveríamos a hacer lo que el señor en cuestión nos solicita, previendo también las consecuencias implicadas, pues nuestra vida, nuestros sueños y anhelos, pueden quedar confinados en la celda de una fría cárcel.



Ahora bien, por su parte, ¿qué puede perder el robot en cuestión? ¿Acaso sabe de qué se trata anhelar una vida próspera?, ¿tiene sensaciones asociadas la consecución de logros profesionales?, ¿desarrolló algún sentido de pertenencia al grupo de cuidadores en el hogar geriátrico donde colaboraba?, ¿imaginó que podía perder la oportunidad de cuidar adultos mayores?, ¿comprende el sufrimiento de la soledad en la vejez?, o ¿pudo desistir de ayudarlo a dicho señor a terminar con ese sufrimiento, al considerar que es un crimen, implica severos castigos y además causa más sufrimiento?

La intención de plantear estas preguntas es detallar algunas de las cuestiones personales y culturales implicadas en una situación semejante, asumiendo además que se tiene un cuerpo situado en un lugar donde se interactúa con otros, bien sea personas o robots sociales. Por ello, se requiere de modelos computacionales de cognición individual y cultural para que un robot humanoide, sobre ruedas o con piernas, adquiera conocimiento general sobre el mundo, pero sobre todo moral, entendido como un tipo particular de conocimiento orientado a la conducta en sociedad, mediante su inmersión en el entorno socio-cultural con el que interactuará. Por esta razón, estos artefactos deberán ser educados por los individuos de dicha sociedad.

Los robots para el cuidado de adultos mayores en su propia casa es una de las tecnologías de mayor impacto en la sociedad contemporánea,

## Moralidad Artificial

Actualmente la roboética se preocupa, en primera instancia, por las cuestiones e implicaciones éticas en el desarrollo de robots que, de algún modo, interactúan con seres humanos y

lo que requiere hacer un esfuerzo para ofrecer robots moralmente seguros (Allen y Wallach, 2012). La población de adultos mayores está creciendo a nivel mundial; no obstante, la búsqueda de su bienestar parece estar relegada. En Bogotá se suicida un adulto mayor a diario, según Medicina Legal en 2017, y 4 adultos mayores son abandonados a diario, según la Asociación Colombiana de Geriatria. Esto da muestras de por qué Colombia se ubica en el peor lugar para envejecer en Latino América (RCN, 2018), tanto por el grave estado afectivo de esta población, como por sus problemáticas económicas y sus difíciles condiciones médicas.

Así, al verse reducido el número de seres humanos dispuestos a cuidar a sus padres y abuelos, los robots socialmente asistentes se han comenzado a ver como un servicio de asistencia permanente, e incluso como compañía alternativa, pero aún deben garantizar cierta naturalidad (Tapus et al., 2007) y plena confiabilidad (Espingardeiro, 2014). Entre otras, esta situación social muestra la pertinencia de usar robots socialmente asistenciales con autonomía moral, pues, además de los adultos mayores, existen muchas otras personas que en algún sentido no cuentan con sus congéneres. Con el desarrollo de este tipo de robots se da también la posibilidad de llegar a comprender mejor nuestra propia naturaleza moral (Allen y Wallach, 2012), la posibilidad de ver nuestra mecanicidad psicológica (Krishnamurti, 1984), y quizás el origen de la maldad, en una especie de espejo tecnológico.

su difusión en la sociedad. En segunda instancia, establece una ética robótica, es decir, los códigos de comportamiento que resulten necesarios para evidenciar funcionamientos aceptablemente

éticos en robots que interactúen con personas. En última instancia, considera la posible existencia de robots que sean capaces de escoger y ejecutar acciones libremente, que comprendan las consecuencias de esas acciones en la interacción con personas y otros agentes en una sociedad, y a los que se les pueda, además, atribuir responsabilidades y derechos (Veruggio, Solis y Van der Loos, 2011).

Se puede afirmar que en los dos últimos escenarios se procura establecer la base de una moralidad artificial, con el fin de lograr robots autónomos y moralmente competentes (Malle, 2016). Pero en ello es urgente que humanistas e ingenieros aúnen esfuerzos direccionados a este objetivo. Tratándose de artefactos socio-técnicos (Vermaas, et al. 2011), el reconocimiento de estos robots como agentes morales y pacientes morales (Rowlands, 2012) resultaría de su interacción con los individuos de la sociedad en la que presten sus servicios.

Puede afirmarse que los dos escenarios anteriormente mencionados también corresponden con los dos enfoques que definen el desarrollo de agentes morales artificiales: descendente y ascendente (Allen y Wallach, 2012). El enfoque descendente se basa en conceptos y teorías éticas, tales como el imperativo categórico, la deontología, el consecuencialismo o el utilitarismo,

en procura de establecer reglas generales para tener un comportamiento moral. El enfoque ascendente, por su parte, se inspira en los procesos de la evolución natural y el aprendizaje, en los que el sentido cooperativo y auto-organizativo han permitido la emergencia, difusión y evolución de comportamientos morales, en sociedades humanas y animales (Bekoff y Pierce, 2009) y hasta la posible existencia de cierta moralidad innata (Hauser, 2006; Allen y Wallach, 2012).

Davenport (2014) establece que los agentes morales artificiales deben ser sociales y cooperativos, es decir, deben considerar los intereses de otros agentes morales y las consecuencias de sus propias acciones, para tener el beneficio de sobrevivir al pertenecer a cierto grupo social. También reconoce a los perros de búsqueda y rescate como agentes y pacientes morales, pues, pese a su “inconsciencia”, sus acciones aportan beneficios colectivos y por ello son recompensados. Esto mismo puede considerarse, por ejemplo, para un robot al cuidado de adultos mayores, pues, asumiendo que no tiene la voluntad de escoger libre y responsablemente sus actuaciones, además de su incapacidad para comprender el efecto de su “buen” comportamiento, bajo el argumento de la habitación china (Copeland, 1993), lo realmente relevante de su servicio es su aporte al bienestar de una persona vulnerable que lo necesita permanentemente.

## En la habitación china

El argumento de la habitación china desvirtúa que pueda haber una máquina capaz de comprender sus acciones más allá del simple hecho de entregar un “buen” resultado, mediante la manipulación de símbolos (Copeland, 1993). Sin embargo, en el experimento mental allí planteado, el agente manipulador de símbolos no

comprende por el hecho de no conocer el idioma chino. De esta manera, vale la pena considerar cómo podemos comprender, mediante el uso social de un lenguaje con el cual nos referimos a cosas del mundo exterior y a nuestros estados mentales, tales como ideas, anhelos, sentimientos, etc.

En esta reflexión no se pretende plantear algún contra-argumento al respecto. Tan solo continuar el experimento mental, imaginando posibilidades que puedan ser eventualmente implementadas en un sistema convencional de cómputo. Adicionalmente, tampoco se pretende inducir la idea de “conciencia artificial”, ni cuestionar si la inteligencia artificial, emocionalidad artificial, o moralidad artificial, puedan ser legítimas, o no, pues, independientemente del debate académico o coloquial al respecto, un robot con ciertas características, en circunstancias particulares, puede provocar la muerte a una persona. Frente a ello, se puede preguntar si la máquina tenía la autonomía suficiente para evitarlo, o incluso, la intención de hacerlo, auto-justificándose de alguna manera, o con base en algún motivo. En consecuencia, acá se indaga sobre la responsabilidad involucrada en tales artefactos.

## La Mecánica Moral

La cuestión de si la moral depende o no de la biología merece ser revisada con cuidado, pues la cooperación entre los individuos de las sociedades humanas primitivas permitió la emergencia y la difusión de nuestros comportamientos morales; pero también lo hacen los genes que nos dotan con un cerebro capaz de realizar dicha cooperación, semejante a como se propagan los genes relacionados con la asimilación de lactosa en sociedades con tradiciones culturales vacuñas (Feldman y Laland, 1996).

Hauser (2006) desarrolla el concepto de un “órgano moral” y una gramática moral universal, la cual plantea una estrecha relación entre la biología y la moral, lo que, de ser cierto, puede desvirtuar el computacionalismo considerado en Davenport (2014). Esta dificultad no es

Quizás no pueda salirse de la habitación china, incluso siendo un ser humano con las limitaciones propias del lenguaje con el cual llegamos a comprender tan solo algunos aspectos del mundo. Es decir, no todas las personas comprendemos la cuántica, ni la teosofía, por mencionar solo un par de contextos en donde muchos sentiríamos que nos hablan en chino. Sin embargo, es posible emprender el largo camino de la comprensión, mediante la adquisición del respectivo lenguaje, y demás tipos de experiencias implicadas. En robótica se tiene el iCub (Robohub, 2015) como exponente de la cognición basada en la experiencia, que involucra tener un cuerpo móvil con visión, audición y sensibilidad táctil, en interacción con el mundo y con otras mentes incorporadas allí también.

desalentadora y merece ser explorada mediante alguna técnica de computación evolutiva, pues, en su calidad de metaheurísticas poblaciones, sirven para modelar y simular sistemas complejos, con el fin de “juzgar, probar y validar [...] hipótesis, conjeturas y teorías sobre los sistemas naturales” (Maldonado y Gómez-Cruz, 2010), en este caso, la naturaleza biológica y cultural de la moral. Considerando dicha dificultad, y el uso de esas técnicas, acá se considera que el enfoque más apropiado para un agente moral es el ascendente, teniendo en mente diseñar un sistema cognitivo para el aprendizaje y libre aceptación de las normas morales vigentes en una sociedad.

Por otra parte, los sistemas basados en redes neuronales artificiales ofrecen la posibilidad de solucionar la dificultad de establecer una cantidad

finita de reglas morales para atender una infinidad de posibles situaciones con implicación moral. Esto se debe a que son capaces de aprender prototipos a partir de ejemplos del tipo de comportamiento idealmente deseado (Clark, 1996); a la par que son capaces de generalizar dichos prototipos en situaciones semejantes a las que se usaron como ejemplo en el aprendizaje. Así, un sistema basado en red neuronal dispone del saber-cómo necesario para alcanzar, en términos aristotélicos, la excelencia moral y la sabiduría práctica, para conocer por inducción el proceder “correcto”, desearlo mediante el hábito, e intuir el sentido de lo “correcto” con cierto sentido común (DeMoss, 1998).

Entre las metaheurísticas poblacionales se tiene también a los sistemas multi-agente, los cuales también tienen ciertas características propias del paradigma conexionista, pero en lugar de neuronas dispone de agentes sociales y en lugar de conexiones sinápticas tiene vínculos entre ellos.

## Replicadores Culturales

Según Dawkins (1976), los memes fueron originalmente conceptualizados como ciertas unidades de transmisión cultural, supuestamente capaces de “saltar” entre cerebros mediante imitación. Dicha concepción considera que los memes pueden ser canciones, recetas de cocina, protocolos de urbanidad, tradiciones religiosas o artefactos tecnológicos. Sin embargo, Aunger (2002) ha considerado que esa diversidad memética viola un principio mediante el cual un replicador solo puede estar hecho de un único sustrato material. Por lo tanto, en el caso de existir realmente, los memes serían exclusivamente neuro-memes confinados en el cerebro.

De esta manera, un sistema multi-agente tiene la capacidad de lidiar con problemas complejos, al establecer protocolos de comunicación entre ellos para la cooperación, coordinación y la resolución de conflictos ante la escasez de recursos disponibles (Ferber, 1999).

Normalmente cuando se visualiza a un sistema basado en inteligencia artificial resolviendo un problema, se piensa en un único agente dotado con algunas capacidades cognitivas. Pero Minsky (1988) considera que la inteligencia resulta de la interacción entre muchos agentes simples. Adicionalmente, casi nunca se observa al agente “quien” resuelve un problema, como parte integral de una sociedad, y aún menos, se le visualiza compuesto por alguna especie de micro-sociedad mental de agentes. Aun así, Sterpin (2015) exploró la utilidad cognitiva de ello, para un robot móvil, mediante una comunidad de neuro-memes.

Aunger (2002) conjetura que la replicación de los neuro-memes debió originarse intra-cerebralmente y que cuando los neuro-memes lograron emitir señales sociales estos se replicaron inter-cerebralmente. Por esta razón, los memes no “saltan” entre cerebros, pues dichas señales no transportan el material cerebral del cual están hechos, sino solo la información necesaria para que la replicación neuro-memética suceda en un nuevo anfitrión con cerebro, mientras afuera parece ser un individuo “imitando” a otro.

En la hipótesis neuro-memética sobre la evolución cultural es en donde encuentro los conceptos neodarwinianos esenciales para tener un

mecanismo de aprendizaje cultural, orientado a la adquisición de los prototipos de comportamiento moral en cierta sociedad. Para ello, en términos técnicos, deben definirse microagentes, capaces de replicación neuro-memética, compitiendo y cooperando en el interior de un macroagente, dotándolo así con el conocimiento necesario para abordar situaciones de carácter moral.

Como el lenguaje sirve de apoyo para especificar ciertos contextos de acción y para resolver

## Conclusiones

En respuesta a la ilusión implicada en aumentar la autonomía de los robots en cuestión para eludir la responsabilidad como diseñador (Allen y Wallach, 2012), declaro acá que mi propia intención en el desarrollo de este tipo de artefactos no es renunciar a la responsabilidad de los eventuales daños que provoque dicha tecnología, e insisto en mi propio compromiso frente al estudio de la naturaleza moral en las sociedades humanas.

En el modelamiento y la simulación del comportamiento moral, sería prudente considerar la noción de maldad que subyace al sadismo humano, desde el *bullying* escolar hasta el comportamiento criminal. Si resultara que el modelo computacional no permitiera evitar un comportamiento con tendencia criminal, es responsable establecer que no debe usarse en robots, sino solo en el estudio de esa problemática social y en un escenario completamente controlado y aislado. Considerando lo sucedido con Tay, el chatbot que se dispuso en Twitter para aprender de las interacciones con el público en general, pero debió quitarse de allí porque en menos de

problemas en cooperación con otros (Clark, 1996), el modelo de aprendizaje cultural debe poder comunicarse en el idioma propio de la sociedad en donde preste sus servicios. En virtud de esto, y asumiendo que debe aprenderse como un niño, debe indagarse en qué medida el aprendizaje del lenguaje socialmente compartido apoya la asimilación de las características culturales “incorrectas”, tal como lo analiza Tolton (2013), con respecto a la propagación y aceptación social de la violencia intrafamiliar.

24 horas había aprendido a expresar xenofobia (BBC, 2016); un robot móvil con visión, audición, brazos y manos, no puede quedar expuesto a la opción de ser convencido para aniquilar a ciertos tipos de personas.

Cada día, todo ser humano es expuesto a la xenofobia, y demás fanatismos, pero según su propia naturaleza, además de su educación familiar y escolar, solo algunos son convencidos de participar en actividades terroristas, o de realizar algún otro tipo de crimen. Así mismo, Tay es un gran ejemplo del peligro involucrado, pues es como colocar convictos a enseñar sus ideas en escuelas infantiles. La sociedad misma no permite que los niños queden expuestos a peligros así, y durante cierta cantidad de tiempo son educados, antes de concederles la autonomía propia de los adultos. Así mismo puede concebirse la educación necesaria para un robot con la autonomía moral suficiente para ser considerado como un agente moral. Tal vez no consuma tiempo en años, pues la naturaleza misma de la computación en entornos virtuales quizás pueda acelerar dicho proceso educativo.



En un reciente experimento con robots antropomorfos, cuyo controlador fue combinado con cierto módulo ético, se observó que un robot con ese tipo de arquitectura puede cambiar, de altruista a egoísta, sin cambiar nada más que cierto valor numérico, de positivo a negativo, en su maquinaria cognitiva (Vanderelst y Winfield, 2016). Al respecto, me atrevo a conjeturar,

por ahora, que el origen de la maldad está en el egoísmo, y desde el desarrollo del perceptrón autosupervisado, ha sido mi responsabilidad concertar recursos filosóficos, científicos y tecnológicos, en lograr comprender el egoísmo, desde su papel en los mecanismos evolutivos (Dawkins, 1976), hasta sus implicaciones psicológicas y sociales en la vida cotidiana.

## Referencias

- Allen, C. y Wallach, W. (2012). Moral machines: contradiction in terms or abdication of human responsibility. En P. Lin, G. Bekey y K. Abney (eds.) *Robot Ethics: The Ethical and Social Implications of Robotics* (55-68), Cambridge: MIT Press.
- Allen, C., Wallach, W. y Smith, I. (2006). Why machine ethics? *IEEE Intelligent Systems* 21(4), 12-17. doi: 10.1109/MIS.2006.83.
- Aunger, R. (2002). *The electric meme: A new theory of how we think*. Free Press.
- BBC News, (2016). Tay, la robot racista y xenófoba de Microsoft. *BBC News*. Recuperado de [https://www.bbc.com/mundo/noticias/2016/03/160325\\_tecnologia\\_microsoft\\_tay\\_bot\\_adolescente\\_inteligencia\\_artificial\\_racista\\_xenofoba\\_lb](https://www.bbc.com/mundo/noticias/2016/03/160325_tecnologia_microsoft_tay_bot_adolescente_inteligencia_artificial_racista_xenofoba_lb)
- Bekoff, M. y Pierce, J. (2009). Wild Justice: Honor and Fairness among Beasts at Play. *American Journal of Play*, 1(4), 451-475.
- Chen, X., Ong, Y. S., Lim, M. H. y Tan, K. C. (2011). A multi-facet survey on memetic computation, *IEEE Transactions on Evolutionary Computation*, 15(5), 591-607.
- Clark, A. (1996). Connectionism, moral cognition, and collaborative problem solving. En L. May, A. Clark y M. Friedman (eds.) *Mind and morals: Essays on cognitive science and ethics*, 109-127, MIT Press.
- Copeland, B. J. (1993). *Artificial Intelligence. A Philosophical Introduction*. Wiley-Blackwell.
- Davenport, D. (2014). Moral mechanisms. *Philosophy & Technology*, 27(1), 47-60.
- Dawkins, R. (1976). *The Selfish Gene*. Oxford: Oxford University Press.
- DeMoss, D. (1998). Aristotle, connectionism, and the morally excellent brain. *Proceedings of the 20th World Congress of Philosophy*, 19, 13-20.

- Espingardeiro, A. (2014). A roboethics framework for the development and introduction of social assistive robots in elderly care (tesis doctoral), Universidad de Salford, Manchester, UK. Recuperado de [http://usir.salford.ac.uk/id/eprint/30815/1/A\\_Roboethics\\_Framework\\_for\\_the\\_Development\\_and\\_Introduction\\_of\\_Social\\_Assistive\\_Robots\\_in\\_Elderly\\_Care.pdf](http://usir.salford.ac.uk/id/eprint/30815/1/A_Roboethics_Framework_for_the_Development_and_Introduction_of_Social_Assistive_Robots_in_Elderly_Care.pdf)
- Feldman, M. y Laland, K. (1996). Gene-culture coevolutionary theory. *Trends in ecology & evolution*, 11(11), 453-457.
- Ferber, J. (1999). *Multi-agent systems: An introduction to distributed artificial intelligence*. Addison-Wesley Longman.
- Gabora, L. (1995). Meme and Variations: A computer model of cultural evolution. In L. Nadel y D. Stein (eds.), *1993 Lectures in complex systems* (pp. 471-486). Boston: Addison-Wesley.
- Gabora, L. (1995). Meme and variations: A computational model of cultural evolution. En *1993 Lectures in complex systems*, 471-485. Addison Wesley.
- Hauser, M. (2006). *Moral minds: How nature designed our universal sense of right and wrong*. Ecco/HarperCollins Publishers.
- Krishnamurti, J. (1984). *La madeja del pensamiento*. España: Edhasa.
- Krueger, J. (2013). Empathy. In B. Kaldis (ed.), *Encyclopedia of Philosophy and the Social Sciences*. Sage Publications.
- Maldonado, C. y Gómez-Cruz, N. (2010). Modelamiento y simulación de sistemas complejos. *Borradores de Investigación: Serie documentos Administración*, 66,21.
- Malle, B. F. (2016). Integrating robot ethics and machine morality: the study and design of moral competence in robots. *Ethics and Information Technology*, 18(4), 243-256.
- Marriott, C. y Chebib, J. (2016). Modeling the evolution of gene-culture divergence. *Proceedings of the Artificial Life Conference*, 500-507.
- Minsky, M. (1988). *Society of mind*. Simon and Schuster.
- Robohub, (2015). Robot: iCub: This humanoid helps us study the brain. National Geographic. *National Geographic* Recuperado de <https://www.nationalgeographic.org/article/robot-icub/>
- RCN Noticias. (2018, 16 de septiembre). Uno de cada diez adultos mayores en Colombia está en el abandono. *Noticias RCN*. Recuperado de <https://noticias.canalrcn.com/nacional-pais/uno-cada-diez-adultos-mayores-colombia-esta-el-abandono>

- 
- Rowlands, M. (2012). *Can animals be moral?* Oxford University Press.
- Sterpin, D. (2011). Perceptrón auto-supervisado: Una red neuronal artificial capaz de replicación memética. *Revista Educación en Ingeniería*, 6(12), 90-101. .
- Sterpin, D. (2015). Cognición imitativa para un robot mediante una comunidad de replicadores neuro-meméticos. *Revista Tecnura*, 19(44), 15-32, Universidad Distrital Francisco José de Caldas.
- Stueber, K. (2013). Empathy. En E. Zalta (ed.) *The Stanford Encyclopedia of Philosophy* (Spring 2018 Edition), Recuperado de <https://plato.stanford.edu/archives/spr2018/entries/empathy/>
- Tapus, A., Mataric, M. y Scassellati, B. (2007). Socially assistive robotics. *IEEE Robotics and Automation Magazine*, 14(1), 35-42.
- Tolton, L. (2013). "Porque te quiero te aporreo": Los refranes como estrategia discursiva para la legitimación del maltrato conyugal. *Discurso & Sociedad*, 7(2), 310-337.
- Vanderelst, D. y Winfield, A. (2016). The dark side of ethical robots. arXiv preprint 1606.02583. Recuperado de <https://arxiv.org/pdf/1606.02583v1.pdf>
- Vermaas, P., Kroes, P., van de Poel, I., Franssen, M. y Houkes, W. (2011). A philosophy of technology: from technical artefacts to sociotechnical systems. *Synthesis Lectures on Engineers, Technology, and Society*, 14, 1-134.
- Veruggio, G., Solis, J. y Van der Loos, M. (2011). Roboethics: Ethics applied to robotics. *IEEE Robotics Automation Magazine*, 18, 21-22.
- Wallach, W. (2008). Implementing moral decision-making faculties in computers and robots. *AI & Society*, 22(4), 463-475.